

## CLAIMS

We claim:

1. A method of determining whether a set of nucleotides is within a first nucleic acid sequence, the method comprising: receiving a first and a second nucleotide of a second  
5 nucleic acid sequence, the second nucleotide being a nucleotide after the first nucleotide; combining said first and second nucleotide into a sequential set; and comparing the first set of nucleotides to a first nucleic acid sequence to determine whether the first set of sequential nucleotides is within the first nucleic acid sequence.
2. The method of claim 1, wherein if the first set of nucleotides is not within the  
10 first nucleic acid sequence, storing said first set as a unit in a database for the second nucleic acid sequence.
3. The method of claim 1, wherein if the first set of nucleotides is within the first nucleic acid sequence, receiving a third nucleotide of the second nucleic acid sequence, the third nucleotide being a nucleotide after the second nucleotide.
- 15 4. The method of claim 3, further comprising: combining the first set of nucleotides with the third nucleotide to make a second sequential set.
5. The method of claim 4, further comprising: comparing the second set of nucleotides to a first nucleic acid sequence to determine whether the second set of sequential nucleotides is within the first nucleic acid sequence.
- 20 6. The method of claim 5, wherein if the second set of nucleotides is not within the first nucleic acid sequence, storing said second set as a unit in a database for the second nucleic acid sequence.
7. The method of claim 6, further comprising: determining the sum of all units stored for the second nucleic acid sequence.
- 25 8. The method of claim 7, further comprising: determining the difference between total number of units stored for a first nucleic acid sequence and the total number of units stored for the second nucleic acid sequence.

9. The method of claim 8, further comprising: utilizing the difference to determine the distance between the first nucleic acid sequence and the second nucleic acid sequence.

10. A computer readable medium comprising the method of claim 1.

5 11. A method of creating a database of nucleotide units for a first nucleic acid sequence, the method comprising: receiving a first nucleotide of a first nucleic acid sequence; determining whether the first nucleotide has been stored in a database as a unit for the first nucleic acid sequence; and if not, storing the first nucleotide as a unit for the first nucleic acid sequence.

10 12. The method of claim 11, wherein if the first nucleotide has been stored in a database as a unit for the first nucleic acid sequence, receiving a second nucleotide of the first nucleic acid sequence, the second nucleotide being a nucleotide after the first nucleotide.

13. The method of claim 12, further comprising: combining the first and second nucleotides into a sequential set.

15 14. The method of claim 13, further comprising: determining whether sequential set has been stored in a database as a unit for the first nucleic acid sequence.

15. The method of claim 14, wherein if the sequential set has been stored, receiving a third nucleotide of the first nucleic acid sequence, the third nucleotide being the next sequential nucleotide after the second nucleotide.

20 16. The method of claim 14, wherein if the sequential set has not been stored, storing said set as a unit for the first nucleic acid sequence.

17. The method of claim 16, further comprising: determining the sum of all units stored for the first nucleic acid sequence.

18. A computer readable medium comprising the method of claim 11.

19. A system for determining whether a set of nucleotides is within a first nucleic acid sequence, the system comprising: a receiving component for receiving a first and a second nucleotide of a second nucleic acid sequence, the second nucleotide being a nucleotide after the first nucleotide; a combining component for combining said first and second nucleotide into a sequential set; and a comparing component for comparing the first set of nucleotides to a first nucleic acid sequence to determine whether the first set of sequential nucleotides is within the first nucleic acid sequence.

20. The system of claim 19, further comprising: a storing component for storing said first set as a unit in a database for the second nucleic acid sequence if the first set of nucleotides is not within the first nucleic acid sequence.

21. The system of claim 20, comprising: a second receiving module for receiving a third nucleotide of the second nucleic acid sequence if it is determined that the first set of nucleotides is within the first nucleic acid sequence.

22. A method of determining the distance between two nucleic acid sequences, the method comprising: determining the number of words in a first nucleic acid sequence; combining the first sequence with a second nucleic acid sequence to make a combined nucleic acid sequence; determining the number of words in the combined nucleic acid sequence; and determining the difference between the number of words in the combined nucleic acid sequence and the first nucleic acid sequence to determine the distance between the first nucleic acid sequence and the second nucleic acid sequence.

23. A computer readable medium comprising the method of claim 22.